DOCUMENT RESUME

ED 188 075 CG 014 472

AUTHOR Surber, Colleen F.
TITLE Effects of Information Reliability in Predicting Task Performance Using Ability and Effort.
INSTITUTION Wisconsin Univ., Madison.
SPONS AGENCY National Science Foundation, Washington, D.C.; Wisconsin Univ., Madison. Graduate School.
PUB DATE May 80
GRANT BNS-7912414
NOTE 15p.; Paper presented at the Annual Meeting of the Midwestern Psychological Association (52nd, St. Louis, MO, May 1-3, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Ability; Adults; *Attribution Theory; *Credibility; *Information Utilization; *Performance; Prediction; *Predictive Measurement; *Reliability; Research Projects; Weighted Scores

ABSTRACT

The examination of credibility effects in predicting achievement, an important step in the study of source credibility effects on attributions, substantiates Birnbaum's findings that variation in the credibility of information can be represented by changes in the weight of the information. Undergraduate subjects (N=65) predicted the performance of hypothetical students on a comprehensive college final exam of medium difficulty, based on IQ scores and study time efforts ratings of varying reliability. The reliability of effort and ability (IQ) information was manipulated to test averaging and multiplying models for differences in prediction. Results indicated that increased reliability for either ability or effort information had greater effects on judged performance. Increased reliability of one type of information lessened the effect of the other type of information. The findings are consistent with an averaging model in which reliability of information influences its weight, but are inconsistent with a multiplying model. (NRB)

EFFECTS OF INFORMATION RELIABILITY IN PREDICTING TASK PERFORMANCE USING ABILITY AND EFFORT

Colleen F. Surber

University of Wisconsin-Madison

Paper presented at the annual convention of the

Midwestern Psychological Association,

St. Louis, MO, May 1980

This research was supported in part by Grant BNS-7912414 from the National Science Foundation, and by a grant from the Graduate School, University of Wisconsin. The author is grateful to Michael H. Birnbaum whose feedback on previous work led directly to the present research. The author would also like to thank Barbara Stevens for her assistance in collecting and analyzing data.

EFFECTS OF INFORMATION RELIABILITY IN PREDICTING TASK PERFORMANCE USING ABILITY AND EFFORT

In the literature on achievement motivation and achievement attributions, a commonly stated hypothesis is that performance is predicted to be a multiplicative function of effort and ability. For example, Heider (1958, p. 83) stated the hypothesis as follows:

> "The personal constituents, namely power and trying, are related as a multiplicative combination, since the effective personal force is zero if either of them is zero. For instance, if a person has the ability, but does not try at all, he will make no progress toward the goal."

Heider here was discussing how people <u>view</u> ability and effort as determining performance (i.e., "naive psychology").

Support for the multiplicative combination rule for judgments of the performance of a hypothetical other has been obtained by Anderson and Butzin (1974) and Kun, Parsons, and Ruble (1974). Anderson and Butzin, however, suggested that an averaging model in which the weights were allowed to vary with the scale values might provide a competing representation of judgments of performance based on ability and effort information. Neither Anderson and Butzin's work nor that of Kun et al. were designed to test the possibility that an averaging model might also accommodate the results. Recent work by Singh, Gupta and Dalal (1979) concluded that in the Indian culture the combination of ability and motivation can be represented by an averaging model. They attributed their results to a difference between the Indian and U.S. cultures. Results reported by Surber with U.S. populations (in press; Note 1) also suggested that the averaging hypothesis is merited further examination.

Previous tests of the averaging model by Singh et al. and Surber relied on the "set size" effect. That is, an averaging model predicts that a piece of information presented alone should have a larger effect than when presented in combination with other information. Unfortunately, there is a class of additive models that are capable of predicting effects of the number of pieces of information (T. Anderson & Birnbaum, 1976; Gollob, Rossman & Abelson, 1973). Thus, the set size effects of Singh et al. and Surber are not definitive. The present work provides a more definitive test of the averaging model by also varying the credibility of information about ability and effort. Recent work on source credibility by Birnbaum (1976; Birnbaum & Stegner, 1979; Birnbaum, Wong & Wong, 1976) provides evidence that variation in the credibility of information can be represented by change in the weight of the information. The present experiment extends Birnbaum's analysis of source credibility effects to predictions of academic performance. In the present experiment, subjects judged the performance of hypothetical students on a comprehensive final exam in a college course. Information about each hypothetical student's intellectual ability was given in terms of an IQ score from one of three different IQ tests described as varying in their reliability. Information about effort was given in terms of estimates of the student's study time for the exam. The information about study time also varied in reliability. Based on this information, subjects predicted the students' performance on the exam.

Although source credibility has long been a topic of interest to social psychologists (e.g., Cohen, 1964; McGuire, 1968), with few exceptions work on source credibility has not addressed predictions of behavior or outcomes. Attribution researchers, however, have recently begun to explore a variety of factors that influence the credibility of information such as "base rates". For example, sample size taken in determining the base rate (Kassin, 1979a), randomness of a sample (Hausen & Donoghue, 1977), and perceived causal relation of the base-rate information to a predicted outcome (Ajzen, 1977; Tversky & Kahneman, 1977) have been examined. As noted by Kassin (1979b) these studies are somewhat atheoretical, though they do provide evidence that the credibility of such variables can be influenced. Birnbaum's model of source credibility has the potential to provide a theoretical umbrella for such phenomena in social attribution. Examination of credibility effects in predicting achievement can be viewed as an important step in laying the groundwork for the study of source credibility effects on attributions.

Models for combining ability and effort

An averaging model for judgments of performance can be written:

$$R = \frac{w_{IQ}s_{IQ} + w_{ST}s_{ST} + w_0 s_0}{w_{IQ} + w_{ST} + w_0}, \qquad (1)$$

where R is the judged performance, w_{IQ} and w_{ST} are weights of IQ and study time information that depend on the reliability of the information, w_0 is the weight of the initial impression, and s_0, s_{IQ} and s_{ST} are the scale values of the initial impression (i.e., expected performance in the absence of any information), the IQ information and the study time information, respectively.

The multiplying model for judgments of performance ($R = s_{IQ}s_{ST}$) makes no provision for variation in the reliability of IQ and study time information. One could propose a kind of weighted multiplying model, however:

$$R = (w_{IQ}s_{IQ})(w_{ST}s_{ST}). \qquad (2)$$

Intuitively, this model can be conceptualized as a two-step integration model in which the subject first combines the weight of each type of information with the value of it (e.g., weight of IQ combined with value of IQ, yields a net impression of the IQ information). Second, the subject combines the IQ impression with the study time impression multiplicatively.

The averaging model of Equation 1 and the multiplying model of Equation 2 both predict that as the reliability of a type of information increases, the effect of that information on the judgment should also increase. For example, the more reliable the IQ information, the greater the predicted effect of IQ. This can be seen in that in both equations the weight multiplies the scale value (e.g., $w_{IQ}s_{IQ}$).

The averaging and multiplying models differ in the predicted effect of the reliability of one type of information on the impact of other information. The averaging model predicts that as the reliability of one type of information increases, the net effect of the other information decreases. For example, as the reliability of IQ increases, the effect of study time on the judgment should decrease. This can be seen by considering the relative weights of IQ and study time. The relative weight of study time, $w'_{ST} = w_{ST}/(w_{ST} + w_{IQ} + w_0)$, will decrease as the value of w_{IQ} increases. In contrast, the multiplying model of Equation 2 predicts that increasing the reliability of one type of information will increase the impact of the other information on the judgment. The predictions of the models differ in other ways as well. For example, Equation 2 predicts a four-way interaction while Equation 1 does not. Equation 2 also predicts a bilinear interaction of the levels of IQ with the levels of study time where Equation 1 does not. In addition, the averaging model predicts that the amount of information presented will influence the impact of each piece of information (the set size effect) while the multiplying model does not.[1]

METHOD

Instructions

Subjects were told that the purpose of the experiment was to examine how people use information about a student's ability and effort to predict performance on an exam. The exam was described as a comprehensive final in a college course that was of medium difficulty.

IQ information. The instructions stated that information about a student's intellectual ability would be given in terms of an IQ score, and that in different cases the IQ score was obtained from test procedures that differed in reliability. The low reliability IQ test scores were described as based on a short written, group administered IQ test taking only 10 minutes. The short IQ test was described as open to many sources of possible error, e.g., lack of attention to the test, luck in guessing correct answers, etc. The instructions also stated that while the short IQ test provides some information about a student's intelligence, it is the most likely to be in error. The medium reliability IQ test scores were described as based on an individually administered test, requiring about an hour. This test was described as more likely to give a good indication of a student's true intelligence because of the larger number of items and the fact that the test is individually administered. The high reliability IQ test scores were described as based on three repeated administrations of the medium reliability IQ test, using a different form of the test each time. The instructions stated that the average of the three scores provided a highly reliable measure of true IQ because of the large variety of test items, administration of the test on three separate days, etc. This procedure was described as producing an IQ score that is "as close as you can get to the student's true IQ."

Study time information. Information about study time was given in terms of how much the student studied for the course compared to others. This information was described as obtained by having students record their amount of studying for various periods of time. Subjects were told to assume that all students reported their study time truthfully. The low reliability study time estimate was described as based on the amount of time the student spent studying for the course for one randomly selected day during the semester. This estimate was described as not a very reliable estimate of overall effort in the course. Factors such as exams in other courses or other activities may have conflicted with the student's study effort on that day. Similarly, a high study time for a single day may not be a good indicator because the day may be atypical. The medium reliability study time estimate was described as based on recorded study time for a whole week during the semester. This procedure was described as more likely to give a reliable indicator of overall study effort than the one day estimate. The high reliability study time estimate was described as based on recorded study time for a whole month during the semester. This procedure was described as the most likely to givr a reliable estimate of the student's overall effort in the course.

Design

There were 144 trials generated by a 3(Reliability of IQ) × 4(Level of IQ) × 3(Reliability of Study Time) × 4(Level of Study Time) factorial design. The levels of IQ were verbally described as well below average, somewhat below average, somewhat above average, and well above average. The 4 levels of Study Time were described in the same way. In addition there were 24 trials generated by a 3(Reliability of IQ) × 4(Level of IQ) design and a 3(Reliability of Study Time) × 4(Level of Study Time) design. These 168 trials were randomly ordered and printed in booklets. The IQ information was printed above the study time information on each trial. The experimental trials were preceded by 22 practice trials, which included some stimuli more extreme than those of the main design (e.g., "extremely above average" or "extremely below average"). Each subject worked at his own pace, with most completing the experiment in approximately one hour.

Rating Scale

The subjects judged performance using integers between 1 and 19, labelled varying from 1 = extremely below average performance, 10 = average, to 19 = extremely above average performance.

Subjects

The subjects were 65 undergraduate students at the University of Wisconsin who participated for extra credit in an introductory psychology course. There were 16 males and 49 females.

RESULTS

Test of the averaging model

The lefthand panel of Figure 1 presents the effects of IQ and IQ reliability on judged performance (averaged across study time and study time reliability). As predicted by both the multiplying and averaging model, as IQ reliability increases the effect of the level of IQ increases. This is also true for the effect of study time and study time reliability which are presented in the righthand panel of Figure 1 (averaged over the levels of IQ and IQ reliability). The IQ × IQ reliability interaction was significant ($F(6, 384) = 120.94$) as was the Study Time × Study Time reliability interaction ($F(6, 384) = 108.59$).

Insert Figures 1 & 2 about here

Figure 2 presents the evidence which distinguishes the averaging from the multiplying model. The lefthand panel of Figure 2 presents the mean judgments of performance as a function of the level of IQ (abscissa) with a different curve for each level of study time reliability. It can be seen that the higher the reliability of study time, the lower the effect of the level of IQ. This finding is predicted by the averaging model, but is contrary to the multiplying model. The Study Time reliability × IQ interaction was significant ($F(6, 384) = 13.41$).

The IQ reliability × Study Time interaction was also significant ($F(6, 384) = 20.41$), and also agrees with the predictions of averaging model (see the right-hand panel of Figure 2). The higher the IQ reliability, the lower the effect of Study Time.

Figure 3 presents the mean judgments of exam performance for the complete 3 × 4 × 3 × 4 design. The 16 points in each panel are the 4 × 4 combinations of IQ and Study Time for one level of IQ reliability combined with one level of Study Time reliability. In each panel, IQ is on the abscissa, and there is a separate curve for each level of study time. The panels in the top row are the mean judgments for the low level of IQ reliability, the middle row for medium IQ reliability, and the bottom row for high IQ reliability. The level of study time reliability increases across the panels from left to right.

The data of Figure 3 can be seen to agree with the predictions of the averaging model. As the level of Study Time reliability increases (as one moves from the left panel to the right panel within each row) the spread of the curves increases. This follows from the fact that the spread of the curves in each panel should be related to the relative weight of study time. Similarly, the effect of IQ reliability can be seen by examining the change in slope within each column. The curves are steeper in the bottom row than in the top row. The effect of IQ reliability can be seen to decrease the effect of study time by noting that within each column of panels, the steeper the slope the <u>smaller</u> the spread of the curves. This follows from the averaging model since increasing the absolute weight of IQ (w_{IQ}) should decrease the relative weight of Study Time ($w_{ST}/(w_{ST} + w_{IQ} + w_0)$). The need for the initial impression in Equation 1 can be seen by examining the panels in the diagonal of Figure 3. In the upper left corner, where the reliability of both cues is low, neither the slope nor spread is very great. In contrast, in the lower right panel where the reliability of both cues is high, both the slope of the curves and the spread of the curves are great. This is predicted nicely by the relative weight averaging model since the effective weight of the initial impression $[w_0/(w_{IQ} + w_{ST} + w_0)]$ should decrease as the values of either w_{IQ} or w_{ST} increase.

Insert Figure 3 about here

The four way interaction predicted by the multiplying model did not materialize ($F(36, 2304) = 1.08$). There was a significant IQ × Study Time interaction ($F(9, 576) = 5.84$), however. This interaction is due to the fact that, averaged over the levels of IQ reliability and Study Time reliability, the curves converge slightly as the level of IQ increases. This interaction differs from Anderson and Butzin's (1974) and Kun et al.'s (1974) results, and is inconsistent with a multiplying model, but is consistent with other findings (Singh et al., 1979; Surber, Note 1). An averaging model can account for such an interaction if the weights are allowed to vary with the scale values (see Birnbaum & Stegner, 1979, for a discussion of configural versus differentially weighted averaging models). There was also a significant interaction of IQ reliability × IQ × Study Time reliability ($F(12, 768) = 2.97$). This interaction was small and appeared to be due to variations in the size of the interaction of Study Time reliability with IQ across levels of IQ reliability. These effects did not appear to be systematic or serious enough to merit further consideration.

Set size effects

An averaging model also predicts that the relative weight of information depends on the number of other pieces of information presented with it. Neither a multiplying nor an additive model predicts effects of the number of sources of information combined. The set size effects predicted by the averaging model can be tested in the present experiment by comparing the effect of Study Time information presented alone with its effect when combined with IQ (and vice versa). Figure 4 presents the mean judgments for the IQ × IQ reliability and the Study Time × Study Time reliability designs. According to the averaging model, the ordinate variation in each panel of Figure 4 should be greater than the ordinate variation in the corresponding panels of Figure 1 (see Birnbaum et al., 1976; Experiment II). This can be shown by a comparison of the relative weights of the information presented alone (e.g., $w_{IQ}/(w_{IQ} + w_0)$) versus in combination with other information ($w_{IQ}/(w_{IQ} + w_{ST} + w_0)$). Comparison of Figure 4 with Figure 1 reveals that these predictions of the relative weight averaging model hold for the present experiment.

The averaging model of Equation 1 was fit to the mean judgments using subroutine STEPIT (Chandler, 1969) to minimize the sum of squared deviations. The weight of the initial impression was set to 1.0. The overall root mean squared error was .290 across the 168 data points. This compares well with the standard errors of the means, which ranged from .115 to .440. Thus, the model predicts the mean judgments within the range of a standard error. The estimated weights of IQ for the three levels of IQ reliability were .397, .793 and 1.040 compared with .317, .564 and .819 for the weights of the three levels of Study Time reliability.

DISCUSSION

The data of the present experiment provide evidence in favor of the averaging model as a representation of the way ability and effort information are combined. In contrast to previous results, this conclusion does not depend on the set size effect, although the set size results agree with the averaging predictions. Thus, it appears that Singh et al. (1979) may have been too hasty in concluding that the Indian and American cultures differ in how they view ability and effort as determining performance.

The fact that the interaction of ability and effort in the present work was not the diverging pattern found by both Anderson and Butzin (1974) and Kun et al. (1974) requires discussion. One possible reason for the difference is that Anderson and Butzin described the tasks as extremely difficult. For example, the instructions for judging graduate school performance stated, "A disturbingly large number of graduate students do not last beyond the first year of study." In the present experiment, the difficulty of the test was purposely described as medium so that the results would be representative of college students' views of performance in college courses. Results similar to the present ones have been obtained in 3 other experiments in which college students judged academic performance (Surber, Note 1, Note 2). Singh et al.'s work, which produced results closely resembling those of the present experiment, included no special instructions pertaining to difficulty. The task was described as performance during the first year engineering curriculum, and the subjects making the judgments were second year engineering students. It may be reasonable to assume that the second-year students regarded the first year curriculum as of medium difficulty, since they are not the ones who flunked out. Based on this analysis, task difficulty may influence the way ability and effort are subjectively combined to predict performance (cf. Kun & Weiner (1973)). This hypothesis has been tested more recently by Surber (Note 2). It is possible that prediction of performance in high difficulty tasks is better represented by a multiplying model. In Surber's (Note 2) experiment, tests of the set size effects conformed to predictions of the averaging model even when bilinear interactions were obtained. Thus, Anderson and Butzin's alternative interpretation of their results as consistent with a differentially weighted averaging model appears to be preferable.

Implications for heuristics of judgment

Recently, Ross (1977) discussed the topic of "attributional biases in prediction," employing a variety of heuristic concepts such as representativeness, availability, anchoring and adjustment, concrete vs abstract information, correlation error, regression error, conservatism and nonconservatism. The approach of the present study suggests an alternative to enumerating judgmental heuristics in predicting outcomes. Most of these heuristics can be re-expressed as predictions of algebraic models of judgment.

The translation of heuristic concepts into algebraic models can be best illustrated by Birnbaum's (1976) numerical prediction task, since it is possible to calculate the optimal statistical solution, allowing evaluation of so-called biases. As pointed out by Birnbaum (1976) an averaging model of source reliability effects can predict judgments that others might describe as conservative, counterconservative, optimal or representative. For example, intuitive predictions that agree with an averaging model are consistent with the notion of representativeness, since the intuitive average of two sources of information seems representative of the information. Similarly, the judgments based on single cues in Birnbaum's (1976) study could be called counterconservative or nonconservative since the regression weights were higher than the optimal weights in this condition. Such overuse of correlated cues has also been called the "regression error" by Ross (1977). Conservatism (Peterson & Beach, 1967) can be seen in the fact that Birnbaum found regression weights for one condition of multiple cue predictions to be smaller than the optimal weights. For another condition of multiple cue predictions Birnbaum found regression weights that were approximately optimal. Some of the findings can also be interpreted as an anchoring effect. When a high reliability source is combined with a low reliability source, the judgments were displaced toward the value of the high reliability information. The subject's judgment in this case could be said to be more firmly anchored at a high reliability value, producing less adjustment. Happily, all the results in Birnbaum's numerical prediction task can be predicted by an averaging model in which the weights of the cues depend on the reliability of the cues. Thus, the model can provide a unifying theoretical framework for predicting when the effects described by the various heuristics will occur.

Since much of the analysis of Birnbaum's numerical prediction task applies analogously to the present experiment, the potential of the model for social attribution is evident. By extending models of source credibility effects to predictions of achievement, the present research suggests a variety of experiments on source credibility in attribution. The most immediate extension would be to examine the effects of information reliability on attributions of ability and effort in an experiment analogous to the present one. For example, the reliability of information about performance and study time might be manipulated while asking for attributions of IQ. A common assumption of attribution theories is that how causes are regarded as determining an effect has an influence on attributions for the effect (Kelley, 1972; Reeder & Brewer, 1979; Zuckerman & Mann, 1979). Based on this assumption (albeit, a questionable one), one might expect to find effects of information reliability on ability and effort attributions that parallel the present results.

Proposing that an averaging model has the potential to describe a variety of source credibility effects in attribution does not mean that it necessarily will be successful. Application of algebraic models of judgment to the effects of variables such as concreteness-abstractness of information, the perceived causal relation of information to outcome, randomness and/or size of samples represented by base-rate information, etc., can serve several purposes. First, it will provide a theoretical context for unifying a set of phenomena in attribution. Second, it will help to discover and define the boundaries of algebraic models of attribution. Third, such research will provide enriched empirical interpretations for the parameters in the models and by doing so should stimulate research into the cognitive processes behind the models (cf. Graesser & Anderson, 1974; Lopes & Ekberg, Note 3; Slovic, Fischoff & Lichtenstein, 1977).

Footnotes

[1]The multiplying model really makes no predictions about the effect of set size on judgment. By assuming that omitted information is replaced by the identity operator, however, the multiplying model predicts that information presented alone should have the same net impact as when combined with other information. Ordinally, the set size predictions of the multiplying model are the same as the ordinal set size predictions of the additive model. The additive model has not been elaborated because it has been compared with the averaging model in detail elsewhere (Birnbaum, 1976; Birnbaum, Wong & Wong, 1976).

Reference Notes

1. Surber, C. F. The nature of cognitive organization in social judgment: Ability, effort and performance. Unpublished manuscript, 1980.

2. Surber, C. F. Effects of task difficulty on judgments of ability, effort and performance. Unpublished manuscript, 1980.

3. Lopes, L., & Ekberg, P. S. Comparison of analog and computational strategies in multiplicative judgment.

References

Ajzen, I. Intuitive theories of events and the effects of base-rate information on prediction. Journal of Personality and Social Psychology, 1977, 35, 303-314.

Anderson, N. H., & Butzin, C. A. Performance = Motivation × Ability: An integration-theoretical analysis. Journal of Personality and Social Psychology, 1974, 30, 598-604.

Anderson, T., & Birnbaum, M. H. Test of an additive model of social inference. Journal of Personality and Social Psychology, 1976, 33, 655-662.

Birnbaum, M. H. Intuitive numerical prediction. American Journal of Psychology, 1976, 89, 417-429.

Birnbaum, M. H., & Stegner, S. E. Source credibility in social judgment: Bias, expertise and the judge's point of view. Journal of Personality and Social Psychology, 1979, 37, 48-74.

Birnbaum, M. H., Wong, R., & Wong, L. K. Combining information from sources that vary in credibility. Memory & Cognition, 1976, 4, 330-336.

Chandler, J. D. Subroutine STEPIT - Finds local minima of a smooth function of several parameters. Behavioral Science, 1969, 14, 81-82.

Cohen, A. R. Attitude change and social influence. New York: Basic Books, 1964.

Dweck, C. S. Achievement. In M. E. Lamb (Ed.), Social and personality development. New York: Holt, Rinehart & Winston, 1978.

Gollob, H. F., Rossman, B. B., & Abelson, R. P. Social inference as a function of the number of instances and consistency of information presented. Journal of Personality and Social Psychology, 1973, 27, 19-33.

Graesser, C., & Anderson, N. H. Cognitive algebra of the equation: Gift size = Generosity × Income. Journal of Experimental Psychology, 1974, 103, 692-699.

Hansen, R. D., & Donoghue, J. M. The power of consensus: Information derived from one's own behavior. Journal of Personality and Social Psychology, 1977, 35, 294-302.

Heider, F. The psychology of interpersonal relations. New York: John Wiley, 1958.

Kassin, S. M. Base rates and prediction: The role of sample size. Personality and Social Psychology Bulletin, 1979, 5, 210-213. (a)

Kassin, S. Consensus information, prediction, and causal attribution: A review of the literature and issues. Journal of Personality and Social Psychology, 1979, 37, 1966-1981. (b)

Kelley, H. H. Causal schemata and the attribution process. New York: General Learning Press, 1972.

Kun, A., Parsons, J., & Ruble, D. Development of integration processes using ability and effort information to predict outcome. Developmental Psychology, 1974, 10, 721-732.

Kun, A., & Weiner, B. Necessary versus sufficient causal schemata for success and failure. Journal of Research in Personality, 1973, 197-207.

McGuire, W. J. The nature of attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), The handbook of social psychology (Vol. 3). Reading, Mass.: Addison-Wesley, 1968.

Nicholls, J. G. Causal attributions and other achievement-related cognitions: Effects of task outcome, attainment value, and sex. Journal of Personality and Social Psychology, 1975, 31, 379-389.

Peterson, C. R., & Beach, L. R. Man as an intuitive statistician. Psychological Bulletin, 1967, 68, 29-46.

Reeder, G. D., & Brewer, M. B. A schematic model of dispositional attribution. Psychological Review, 1979, 86, 61-79.

Ross, L. The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology, Vol. 10. New York: Academic Press, 1977.

Singh, R., Gupta, M., & Dalal, A. K. Cultural difference in attribution of performance: An integration-theoretical analysis. Journal of Personality and Social Psychology, 1979, 37, 1342-1351.

Slovic, P., Fischoff, B., & Lichtenstein, S. Behavioral decision theory. Annual Review of Psychology, 1977, 28, 1-39.

Surber, C. F. The development of reversible operations in judgments of ability, effort and performance. Child Development, in press.

Tversky, A., & Kahneman, D. Causal schemata in judgments under uncertainty. In M. Fishbein (Ed.), Progress in social psychology. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Weiner, B. Achievement attribution and attribution theory. Morristown, N.J.: General Learning Press, 1974.

Zuckerman, M., & Mann, R. W. The other way around. Effects of causal attributions on estimates of consensus, distinctiveness, and consistency. Journal of Experimental Social Psychology, 1979, 15, 582-597.

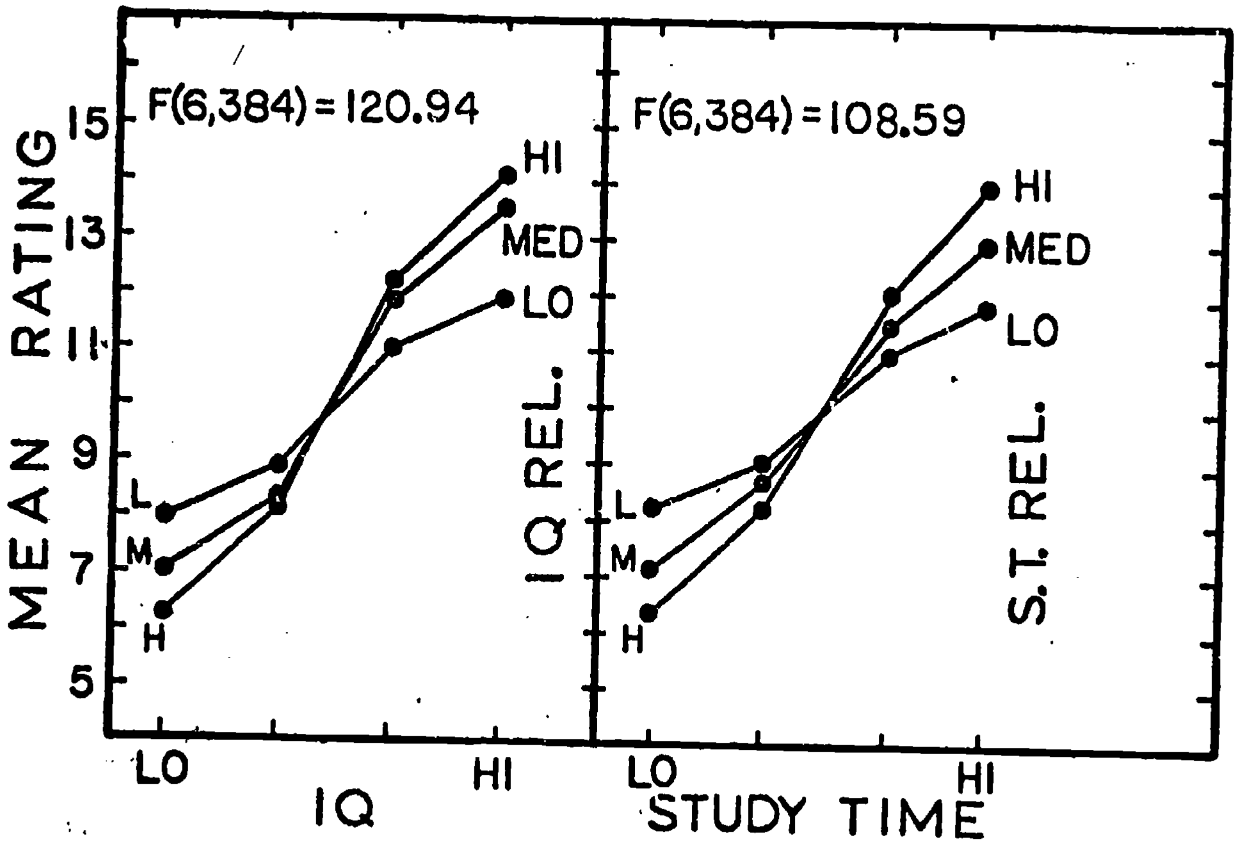
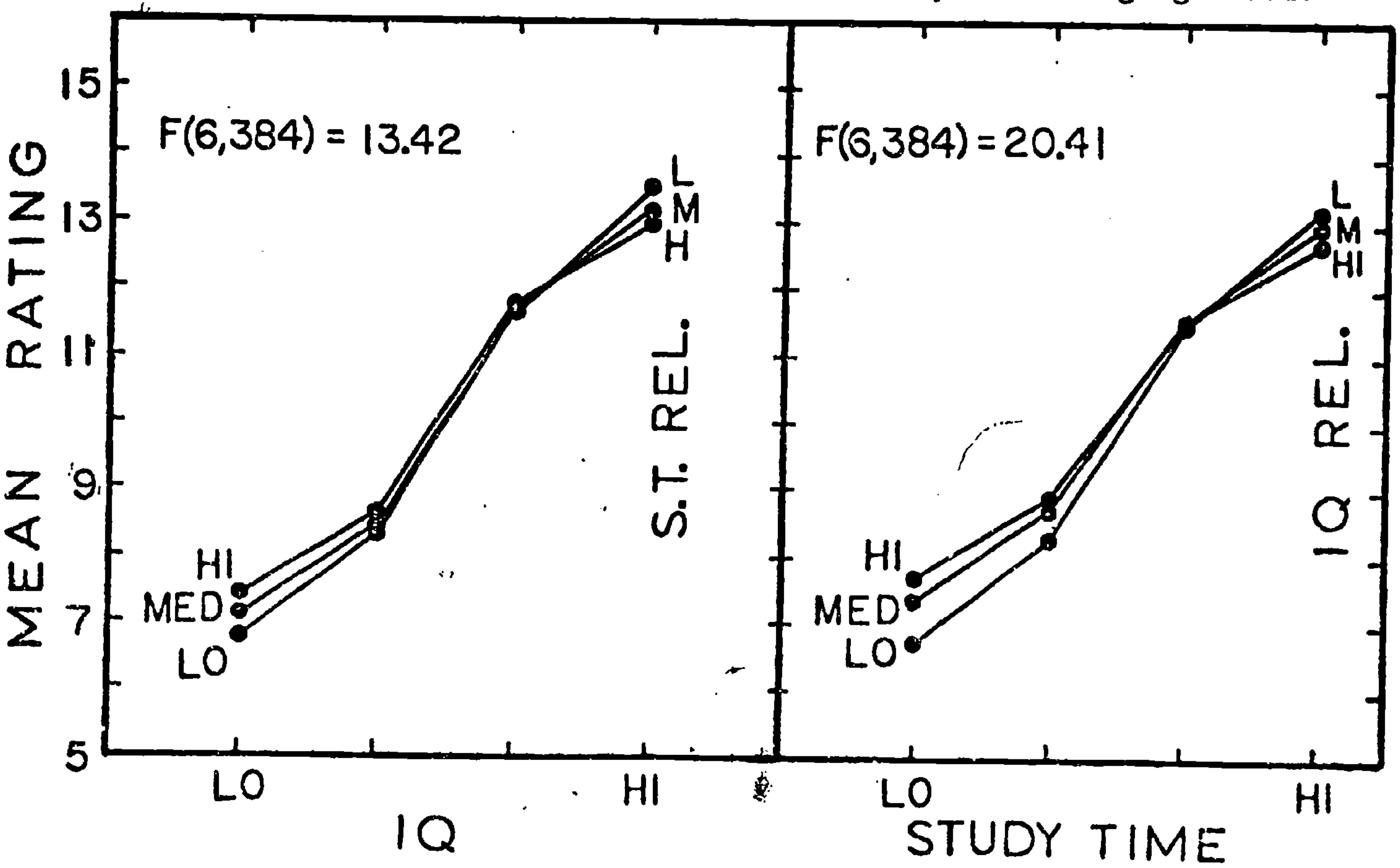


Figure 1. Mean judgments of exam performance as a function of IQ and IQ reliability (left hand panel) and Study Time and Study Time reliability (right hand panel).

Figure 2. Mean judgments of exam performance as a function of IQ and Study Time reliability (left hand panel) and Study Time and IQ reliability (right hand panel). Note that the order of curves in each panel of Figure 2 is the reverse of the order in Figure 1, as predicted by an averaging model.



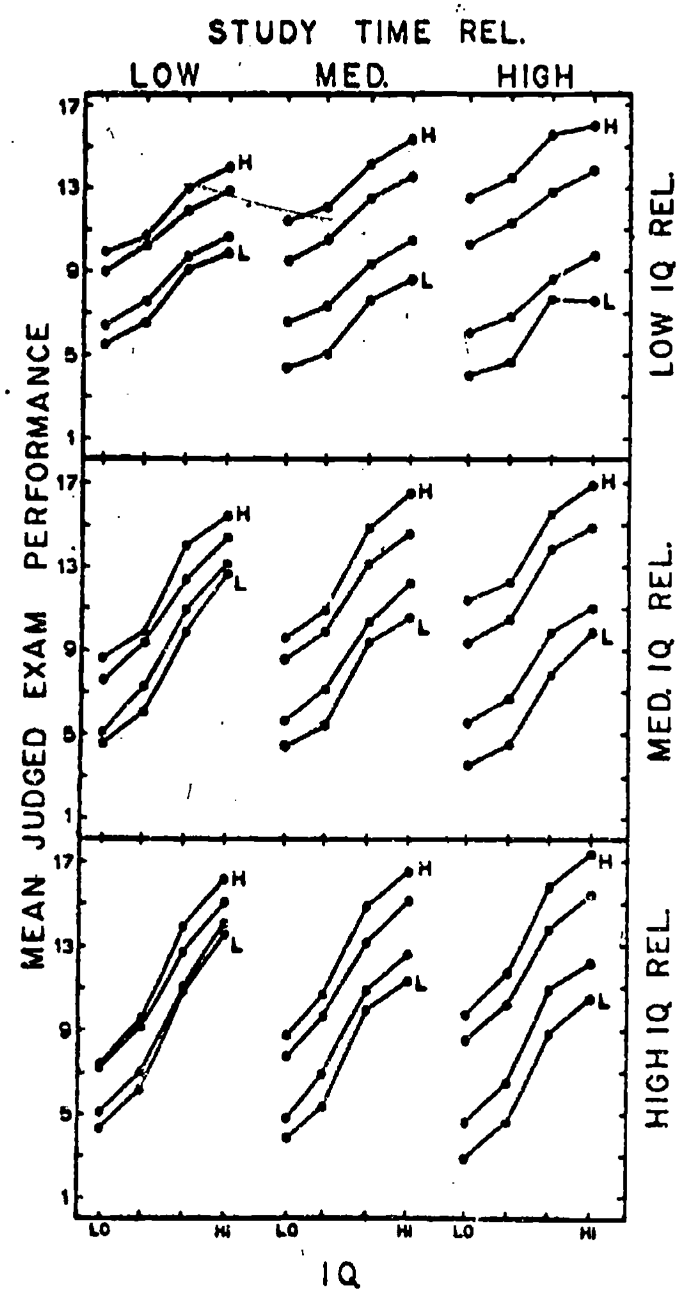


Figure 3. Mean judgments of exam performance as a function of Study Time and IQ information. Each row of panels represents a different IQ reliability; each column of panels represents a different Study Time reliability. In each panel, each solid curve is a different level of Study Time. IQ levels are on the abscissa.

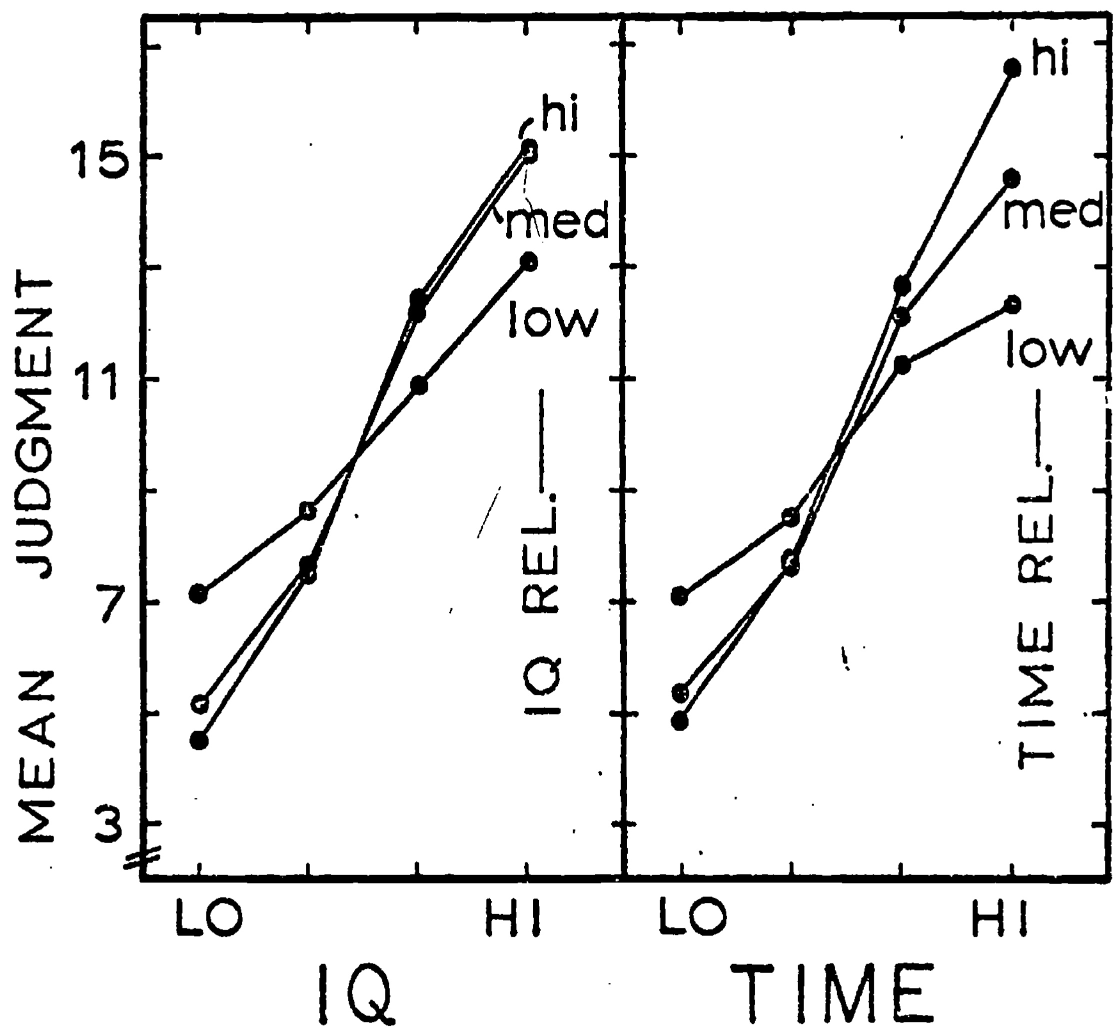


Figure 4. Mean judgments of exam performance for the IQ × IQ reliability design (left panel) and the Study Time × Study Time reliability design (right panel).